(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 July 2001 (12.07.2001)

PCT

(10) International Publication Number
WO 01/49728 A2

(51) International Patent Classification[7]: C07K 14/435

(21) International Application Number: PCT/JP00/09359

(22) International Filing Date:
28 December 2000 (28.12.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

2000-585	6 January 2000 (06.01.2000)	JP
2000-588	6 January 2000 (06.01.2000)	JP
2000-2299	11 January 2000 (11.01.2000)	JP
2000-26862	3 February 2000 (03.02.2000)	JP
2000-58367	3 March 2000 (03.03.2000)	JP

(71) Applicants *(for all designated States except US)*: **PROTEGENE INC.** [JP/JP]; 2-20-3, Naka-cho, Meguro-ku, Tokyo 153-0065 (JP). **SAGAMI CHEMICAL RESEARCH CENTER** [JP/JP]; 4-1, Nishi-Ohnuma 4-chome, Sagamihara-shi, Kanagawa 229-0012 (JP).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **KATO, Seishi** [JP/JP]; 3-46-50, Wakamatsu, Sagamihara-shi, Kanagawa 229-0014 (JP). **KIMURA, Tomoko** [JP/JP]; 715, 2-9-1, Kohoku, Tsuchiura-shi, Ibaraki 300-0032 (JP).

(74) Agents: **AOYAMA, Tamotsu** et al.; AOYAMA & PARTNERS, IMP Building, 3-7, Shiromi 1-chome, Chuo-ku, Osaka-shi, Osaka 540-0001 (JP).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/49728 A2

(54) Title: HUMAN PROTEINS HAVING HYDROPHOBIC DOMAINS AND DNAs ENCODING THESE PROTEINS

(57) Abstract: The present invention provides human proteins having hydrophobic domains, DNAs encoding these proteins, expression vectors for these DNAs, transformed eukaryotic cells expressing these DNAs and antibodies directed to these proteins.

DESCRIPTION

Human Proteins Having Hydrophobic Domains and DNAs Encoding These Proteins

TECHNICAL FIELD

The present invention relates to human proteins having hydrophobic domains, DNAs encoding these proteins, expression vectors for these DNAs, eukaryotic cells expressing these DNAs and antibodies directed to these proteins. The proteins of the present invention can be employed as pharmaceuticals or as antigens for preparing antibodies directed to these proteins. The human cDNAs of the present invention can be utilized as probes for genetic diagnosis and gene sources for gene therapy. Furthermore, the cDNAs can be utilized as gene sources for producing the proteins encoded by these cDNAs in large quantities. Cells into which these genes are introduced to express secretory proteins or membrane proteins in large quantities can be utilized for detection of the corresponding receptors or ligands, screening of novel small molecule pharmaceuticals and the like. The antibodies of the present invention can be utilized for the detection, quantification, purification and the like of the proteins of the present invention.

BACKGROUND ART

Cells secrete many proteins extracellularly. These secretory proteins play important roles in the proliferation control, the differentiation induction, the material transport, the biophylaxis, and the like of the cells. Unlike intracellular proteins, the secretory proteins exert their actions outside the cells. Therefore, they can be administered in the intracorporeal manner such as the injection or the drip, and they possess hidden potentialities as pharmaceuticals. In fact, a number of human secretory proteins such as interferons, interleukins, erythropoietin, thrombolytic agents and the like are currently employed as pharmaceuticals. In addition, secretory proteins other than those described above are undergoing clinical trials for developing their use as pharmaceuticals. It is believed that the human cells produce many unknown secretory proteins. Availability of these secretory proteins as well as genes encoding them is expected to lead to development of novel pharmaceuticals utilizing them.

On the other hand, membrane proteins play important roles, as signal receptors, ion channels, transporters and the like, in the material transport and the signal transduction through the cell membrane. Examples thereof include receptors for various cytokines, ion

channels for the sodium ion, the potassium ion, the chloride ion and the like, transporters for saccharides, amino acids and the like. The genes for many of them have already been cloned. It has been clarified that abnormalities in these membrane proteins are involved in a number of previously cryptogenic diseases. Therefore, discovery of a new membrane protein is expected to lead to elucidation of the causes of many diseases, and isolation of new genes encoding the membrane proteins has been desired.

Heretofore, due to difficulty in the purification from human cells, many of these secretory proteins and membrane proteins have been isolated by genetic approaches. A general method is the so-called expression cloning method, in which a cDNA library is introduced into eukaryotic cells to express cDNAs, and the cells secreting, or expressing on the surface of membrane, the protein having the activity of interest are then screened. However, only genes for proteins with known functions can be cloned by using this method.

In general, a secretory protein or a membrane protein possesses at least one hydrophobic domain within the protein. After synthesis on ribosomes, such domain works as a secretory signal or remains in the phospholipid membrane to be entrapped in the membrane. Accordingly, if the existence of a highly hydrophobic domain is observed in the amino acid sequence of a protein encoded by a cDNA when the

whole base sequence of the full-length cDNA is determined, it is considered that the cDNA encodes a secretory protein or a membrane protein.

OBJECTS OF INVENTION

The main object of the present invention is to provide novel human proteins having hydrophobic domains, DNAs coding for these proteins, expression vectors for these DNAs, transformed eucaryotic cells that are capable of expressing these DNAs and antibodies directed to these proteins.

SUMMARY OF INVENTION

As the result of intensive studies, the present inventors have successfully cloned cDNAs encoding proteins having hydrophobic domains from the human full-length cDNA bank, thereby completing the present invention. Thus, the present invention provides a human protein having hydrophobic domain(s), namely a protein comprising any one of amino acid sequences selected from the group consisting of SEQ ID NOS: 1 to 10, 31 to 40, 61 to 70, 91 to 100 and 121 to 130. Moreover, the present invention provides a DNA encoding said protein, exemplified by a cDNA comprising any one of base sequences selected from the group consisting of SEQ ID NOS: 11 to 30, 41 to 60, 71 to 90, 101 to 120 and 131

to 150, an expression vector that is capable of expressing said DNA by in vitro translation or in eukaryotic cells, a transformed eukaryotic cell that is capable of expressing said DNA and of producing said protein, and an antibody directed to said protein.

This object as well as other objects and advantages of the present invention will become apparent to those skilled in the art from the following description with reference to the accompanying drawings.

BRIEF DESCRIPTION OF DRAWINGS

Figure 1: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP03613.

Figure 2: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP03700.

Figure 3: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP03935.

Figure 4: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10755.

Figure 5: A figure depicting the hydrophobicity/hydrophilicity profile of the protein

encoded by clone HP10760.

Figure 6: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10764.

Figure 7: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10768.

Figure 8: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10769.

Figure 9: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10784.

Figure 10:A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10786.

Figure 11:A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP03727.

Figure 12:A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP03801.

Figure 13:A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP03883.

Figure 14: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP03913.

Figure 15: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10753.

Figure 16: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10758.

Figure 17: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10771.

Figure 18: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10778.

Figure 19: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10781.

Figure 20:A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10785.

Figure 21:A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP03878.

Figure 22:A figure depicting the

hydrophobicity/hydrophilicity profile of the protein encoded by clone HP03884.

Figure 23:A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP03934.

Figure 24: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP03949.

Figure 25: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP03959.

Figure 26: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP03983.

Figure 27: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10745.

Figure 28: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10775.

Figure 29: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10782.

Figure 30:A figure depicting the hydrophobicity/hydrophilicity profile of the protein.

Figure 31:A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP03977.

Figure 32:A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10649.

Figure 33:A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10779.

Figure 34: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10790.

Figure 35: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10793.

Figure 36: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10794.

Figure 37: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10797.

Figure 38: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10798.

Figure 39: A figure depicting the

hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10800.

Figure 40:A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10801.

Figure 41:A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP03596.

Figure 42:A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP03882.

Figure 43:A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP03903.

Figure 44: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP03974.

Figure 45: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP03978.

Figure 46: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10735.

Figure 47: A figure depicting the hydrophobicity/hydrophilicity profile of the protein

encoded by clone HP10750.

Figure 48: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10777.

Figure 49: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10780.

Figure 50:A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10795.

DETAILED DESCRIPTION OF THE INVENTION

The proteins of the present invention can be obtained, for example, by a method for isolating proteins from human organs, cell lines or the like, a method for preparing peptides by the chemical synthesis based on the amino acid sequence of the present invention, or a method for producing proteins by the recombinant DNA technology using the DNAs encoding the hydrophobic domains of the present invention. Among these, the method for producing proteins by the recombinant DNA technology is preferably employed. For example, the proteins can be expressed in vitro by preparing an RNA by in vitro transcription from a vector having the cDNA of the present invention, and then carrying out in vitro translation using this RNA as a

template. Alternatively, incorporation of the translated region into a suitable expression vector by the method known in the art may lead to expression of the encoded protein in large quantities in prokaryotic cells such as *Escherichia coli* and *Bacillus subtilis*, or eukaryotic cells such as yeasts, insect cells and mammalian cells.

In the case where the protein of the present invention is produced by expressing the DNA by in vitro translation, the protein of the present invention can be produced in vitro by incorporating the translated region of this cDNA into a vector having an RNA polymerase promoter, and then adding the vector to an in vitro translation system such as a rabbit reticulocyte lysate or a wheat germ extract, which contains an RNA polymerase corresponding to the promoter. The RNA polymerase promoters are exemplified by T7, T3, SP6 and the like. The vectors containing promoters for these RNA polymerases are exemplified by pKA1, pCDM8, pT3/T7 18, pT7/3 19, pBluescript II and the like. Furthermore, the protein of the present invention can be expressed in the secreted form or the form incorporated in the microsome membrane when a canine pancreas microsome or the like is added to the reaction system.

In the case where the protein of the present invention is produced by expressing the DNA in a microorganism such as *Escherichia coli*, a recombinant

expression vector in which the translated region of the cDNA of the present invention is incorporated into an expression vector having an origin which is capable of replicating in the microorganism, a promoter, a ribosome-binding site, a cDNA-cloning site, a terminator and the like is constructed. After transformation of the host cells with this expression vector, the resulting transformant is cultured. Thus, the protein encoded by the cDNA can be produced in large quantities in the microorganism. In this case, a protein fragment containing any translated region can be obtained by adding an initiation codon and a termination codon in front of and behind the selected translated region and expressing the protein. Alternatively, the protein can be expressed as a fusion protein with another protein. Only the portion of the protein encoded by the cDNA can be obtained by cleaving this fusion protein with a suitable protease. The expression vectors for *Escherichia coli* are exemplified by the pUC series, pBluescript II, the pET expression system, the pGEX expression system and the like.

In the case where the protein of the present invention is produced by expressing the DNA in eukaryotic cells, the protein of the present invention can be produced as a secretory protein, or as a membrane protein on the surface of cell membrane, by incorporating the translated region of the cDNA into an expression vector for eukaryotic

cells that has a promoter, a splicing region, a poly(A) addition site and the like, and then introducing the vector into the eukaryotic cells. The expression vectors are exemplified by pKA1, pED6dpc2, pCDM8, pSVK3, pMSG, pSVL, pBK-CMV, pBK-RSV, EBV vectors, pRS, pYES2 and the like. Examples of eukaryotic cells to be used in general include mammalian cultured cells such as monkey kidney COS7 cells and Chinese hamster ovary CHO cells, budding yeasts, fission yeasts, silkworm cells, and Xenopus oocytes. Any eukaryotic cells may be used as long as they are capable of expressing the proteins of the present invention. The expression vector can be introduced into the eukaryotic cells by using a method known in the art such as the electroporation method, the calcium phosphate method, the liposome method and the DEAE-dextran method.

After the protein of the present invention is expressed in prokaryotic cells or eukaryotic cells, the protein of interest can be isolated and purified from the culture by a combination of separation procedures known in the art. Examples of the separation procedures include treatment with a denaturing agent such as urea or a detergent, sonication, enzymatic digestion, salting-out or solvent precipitation, dialysis, centrifugation, ultrafiltration, gel filtration, SDS-PAGE, isoelectric focusing, ion-exchange chromatography, hydrophobic

chromatography, affinity chromatography and reverse phase chromatography.

The proteins of the present invention also include peptide fragments (of 5 amino acid residues or more) containing any partial amino acid sequences in the amino acid sequences represented by SEQ ID NOS: 1 to 10, 31 to 40, 61 to 70, 91 to 100 and 121 to 130. These peptide fragments can be utilized as antigens for preparation of antibodies. Among the proteins of the present invention, those having the signal sequences are secreted in the form of mature proteins after the signal sequences are removed. Therefore, these mature proteins shall come within the scope of the protein of the present invention. The N-terminal amino acid sequences of the mature proteins can be easily determined by using the method for the determination of cleavage site of a signal sequence [JP-A 8-187100]. Furthermore, some membrane proteins undergo the processing on the cell surface to be converted to the secreted forms. Such proteins or peptides in the secreted forms shall also come within the scope of the protein of the present invention. In the case where sugar chain-binding sites are present in the amino acid sequences of the proteins, expression of the proteins in appropriate eukaryotic cells affords the proteins to which sugar chains are added. Accordingly, such proteins or peptides to which sugar chains are added shall also come

within the scope of the protein of the present invention.

The DNAs of the present invention include all the DNAs encoding the above-mentioned proteins. These DNAs can be obtained by using a method for chemical synthesis, a method for cDNA cloning and the like.

The cDNAs of the present invention can be cloned, for example, from cDNA libraries derived from the human cells. The cDNAs are synthesized by using poly(A)$^+$ RNAs extracted from human cells as templates. The human cells may be cells delivered from the human body, for example, by the operation or may be the cultured cells. The cDNAs can be synthesized by using any method such as the Okayama-Berg method [Okayama, H. and Berg, P., Mol. Cell. Biol. 2: 161-170 (1982)], the Gubler-Hoffman method [Gubler, U. and Hoffman, J., Gene 25: 263-269 (1983)] and the like. However, it is desirable to use the capping method [Kato, S. et al., Gene 150: 243-250 (1994)], as exemplified in Examples, in order to obtain a full-length clone in an effective manner. In addition, commercially available human cDNA libraries can be utilized. The cDNAs of the present invention can be cloned from the cDNA libraries by synthesizing an oligonucleotide on the basis of base sequences of any portion in the cDNA of the present invention and screening the cDNA libraries using this oligonucleotide as a probe for colony or plaque hybridization according to a method known

in the art. In addition, the cDNA fragments of the present invention can be prepared from an mRNA isolated from human cells by the RT-PCR method in which oligonucleotides which hybridize with both termini of the cDNA fragment of interest are synthesized, which are then used as the primers.

The cDNAs of the present invention are characterized in that they comprise any one of the base sequences represented by SEQ ID NOS: 11 to 20, 41 to 50, 71 to 80, 101 to 110 and 131 to 140 or the base sequences represented by SEQ ID NOS: 21 to 30, 51 to 60, 81 to 90, 111 to 120 and 141 to 150. Table 1 summarizes the clone number (HP number), the cells from which the cDNA clone was obtained, the total number of bases of the cDNA, and the number of the amino acid residues of the encoded protein, for each of the cDNAs.

Table 1

Sequence No.	HP No.	Cell	Number of bases	Number of amino acids
1, 11, 21	HP03613	Kidney	2865	578
2, 12, 22	HP03700	Kidney	3323	243
3, 13, 23	HP03935	Kidney	1585	461
4, 14, 24	HP10755	Kidney	2122	647
5, 15, 25	HP10760	Kidney	1775	446
6, 16, 26	HP10764	Kidney	1372	197
7, 17, 27	HP10768	Kidney	2074	540
8, 18, 28	HP10769	Kidney	2252	442
9, 19, 29	HP10784	Kidney	1461	262
10, 20, 30	HP10786	Kidney	1122	152
31, 41, 51	HP03727	Kidney	1617	335
32, 42, 52	HP03801	Umbilical cord blood	1749	208
33, 43, 53	HP03883	Kidney	1402	406
34, 44, 54	HP03913	Kidney	2474	618
35, 45, 55	HP10753	Umbilical cord blood	3296	208
36, 46, 56	HP10758	Kidney	1818	502
37, 47, 57	HP10771	Kidney	1646	336
38, 48, 58	HP10778	Kidney	1416	340
39, 49, 59	HP10781	Kidney	1927	223
40, 50, 60	HP10785	Kidney	1419	309
61, 71, 81	HP03878	Kidney	2016	599
62, 72, 82	HP03884	Kidney	1446	81
63, 73, 83	HP03934	Kidney	2467	654
64, 74, 84	HP03949	Kidney	1450	390
65, 75, 85	HP03959	Kidney	1897	452

invention can be easily obtained by screening the cDNA libraries constructed from the human cell lines or human tissues utilized in the present invention using an oligonucleotide probe synthesized on the basis of the base sequence of the cDNA provided in any one of SEQ ID NOS: 11 to 30, 41 to 60, 71 to 90, 101 to 120 and 131 to 150.

In general, the polymorphism due to the individual differences is frequently observed in human genes. Accordingly, any cDNA in which one or plural nucleotides are added, deleted and/or substituted with other nucleotides in SEQ ID NOS: 11 to 30, 41 to 60, 71 to 90, 101 to 120 and 131 to 150 shall come within the scope of the present invention.

Similarly, any protein in which one or plural amino acids are added, deleted and/or substituted with other amino acids resulting from the above-mentioned changes shall come within the scope of the present invention, as long as the protein possesses the activity of the protein having any one of the amino acid sequences represented by SEQ ID NOS: 1 to 10, 31 to 40, 61 to 70, 91 to 100 and 121 to 130.

The cDNAs of the present invention also include cDNA fragments (of 10 bp or more) containing any partial base sequence in the base sequences represented by SEQ ID NOS: 11 to 20, 41 to 50, 71 to 80, 101 to 110 and 131 to 140 or in the base sequences represented by SEQ ID NOS: 21 to 30, 51 to 60, 81 to 90, 111 to 120 and 141 to 150. Also, DNA

fragments each consisting of a sense strand and an anti-sense strand shall come within this scope. These DNA fragments can be utilized as the probes for the genetic diagnosis.

The antibody of the present invention can be obtained from a serum after immunizing an animal using the protein of the present invention as an antigen. A peptide that is chemically synthesized based on the amino acid sequence of the present invention and a protein expressed in eukaryotic or prokaryotic cells can be used as an antigen. Alternatively, an antibody can be prepared by introducing the above-mentioned expression vector for eukaryotic cells into the muscle or the skin of an animal by injection or by using a gene gun and then collecting a serum therefrom [JP-A 7-313187]. Animals that can be used include a mouse, a rat, a rabbit, a goat, a chicken and the like. A monoclonal antibody directed to the protein of the present invention can be produced by fusing B cells collected from the spleen of the immunized animal with myelomas to generate hybridomas.

In addition to the activities and uses described above, the polynucleotides and proteins of the present invention may exhibit one or more of the uses or biological activities (including those associated with assays cited herein) identified below. Uses or activities described for proteins of the present invention may be provided by

administration or use of such proteins or by administration or use of polynucleotides encoding such proteins (such as, for example, in gene therapies or vectors suitable for introduction of DNA).

Research Uses and Utilities

The polynucleotides provided by the present invention can be used by the research community for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on Southern gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the process of discovering other novel polynucleotides; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination of expression patterns; to raise anti-protein antibodies using DNA immunization techniques;

and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803 (1993)) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The proteins provided by the present invention can similarly be used in assay to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Where the protein binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the protein can be used to identify the other protein with which binding occurs or to identify inhibitors of the binding interaction.

base sequences of the present cDNA has revealed the registration of sequences that shared a homology of 90% or more (for example, Accession No. AA42490) among ESTs. However, since they are partial sequences, it can not be judged whether or not they encode the same protein as the protein of the present invention.

<HP10760> (SEQ ID NOS: 5, 15, and 25)

Determination of the whole base sequence of the cDNA insert of clone HP10760 obtained from cDNA library of human kidney revealed the structure consisting of a 61-bp 5'-untranslated region, a 1341-bp ORF, and a 373-bp 3'-untranslated region. The ORF encodes a protein consisting of 446 amino acid residues and there existed a putative secretory signal at the N-terminus. Figure 5 depicts the hydrophobicity/hydrophilicity profile, obtained by the Kyte-Doolittle method, of the present protein. In vitro translation resulted in formation of a translation product of 48 kDa that was somewhat smaller than the molecular weight of 49,468 predicted from the ORF. In this case, the addition of a microsome led to the formation of a product of 50 kDa. In addition, there exists in the amino acid sequence of this protein two sites at which N-glycosylation may occur (Asn-Ala-Thr at position 144 and Asn-Ile-Ser at position 243). Application of the (-3,-1) rule, a method for predicting the cleavage site of the secretory signal

sequence, allows to expect that the mature protein starts from glutamic acid at position 27.

The search of the protein database using the amino acid sequence of the present protein revealed that the protein was similar to human 25 kDa trypsin inhibitor (Accession No. BAA25066). Table 6 shows the comparison between amino acid sequences of the human protein of the present invention (HP) and human 25 kDa trypsin inhibitor (TI). Therein, the marks of -, *, and . represent a gap, an amino acid residue identical with that of the protein of the present invention, and an amino acid residue similar to that of the protein of the present invention, respectively. The both proteins shared a homology of 33.5% in the intermediate region of 185 amino acid residues.

Table 6

```
HP                    MLHPETSPGRGHLLAVLLALLGTAWAEVWPPQLQEQAPMAG

TI MIAISAVSSALLFSLLCEASTVVLLNSTDSSPPTNNFTDIEAALKAQLDSADIPKARRKR

HP ALNRKESFLLLSLHNRLRSWVQPPAADMRRLDWSDSLAQLAQARAALCGIPTPSLASGLW
   ..... . .*. **..*. * ****.* . *...**. *.*.** * * ... .* *
TI YISQNDMIAILDYHNQVRGKVFPPAANMEYMVWDENLAKSAEAWAATC-IWDHG-PSYLL
```

```
HP RTLQVGWNMQLLPAGLASFVEVVSLWFAEGQRYSHA-AGEC-----AR--NATCTHYTQL
   * *  .  . . . .   *....*. *..* . *. . . .*      *  .. ******.
TI RFLGQN--LSVRTGRYRSILQLVKPWYDEVKDYAFPYPQDCNPRCPMRCFGPMCTHYTQM

HP VWATSSQLGCGRHLCSAGQA--AI---EAF-VCAYSPGGNWEVNGKTIIPYKKGAWCSLC
   *****...**. * * . ..  ..   ... **.*.* ***   *..  *** *. ** *
TI VWATSNRIGCAIHTCQNMNVWGSVWRRAVYLVCNYAPKGNW--IGEA--PYKVGVPCSSC

HP TASVSGCFKAWDHAGGLCEVPRNPCRMSCQNHGRLNISTCHCHCPPGYTGRYCQVRCSLQ
   ..* .*
TI PPSYGGSCTDNLCFPGVTSNYLYWFK
```

The search of the GenBank using the base sequences of the present cDNA has revealed the registration of sequences that shared a homology of 90% or more (for example, Accession No. AI792411) among ESTs. However, since they are partial sequences, it can not be judged whether or not they encode the same protein as the protein of the present invention.

<HP10764> (SEQ ID NOS: 6, 16, and 26)

Determination of the whole base sequence of the cDNA insert of clone HP10764 obtained from cDNA library of human kidney revealed the structure consisting of a 326-bp 5'-untranslated region, a 594-bp ORF, and a 452-bp 3'-untranslated region. The ORF encodes a protein consisting of

500 505 510

Leu Ala Val Gly Ala Ala Phe Ala Ser Ser Trp Lys Thr Gly Leu Ala

515 520 525

Thr Ser Leu Ala Val Phe Cys His Glu Leu Pro His Glu Leu Gly Asp

530 535 540

Phe Ala Ala Leu Leu His Ala Gly Leu Ser Val Arg Gln Ala Leu Leu

545 550 555 560

Leu Asn Leu Ala Ser Ala Leu Thr Ala Phe Ala Gly Leu Tyr Val Ala

565 570 575

Leu Ala Val Gly Val Ser Glu Glu Ser Glu Ala Trp Ile Leu Ala Val

580 585 590

Ala Thr Gly Leu Phe Leu Tyr Val Ala Leu Cys Asp Met Leu Pro Ala

595 600 605

Met Leu Lys Val Arg Asp Pro Arg Pro Trp Leu Leu Phe Leu Leu His

610 615 620

Asn Val Gly Leu Leu Gly Gly Trp Thr Val Leu Leu Leu Leu Ser Leu

625 630 635 640

Tyr Glu Asp Asp Ile Thr Phe

645

<210> 5

<211> 446

<212> PRT

<213> Homo sapiens

<400> 5

Met Leu His Pro Glu Thr Ser Pro Gly Arg Gly His Leu Leu Ala Val
1 5 10 15

Leu Leu Ala Leu Leu Gly Thr Ala Trp Ala Glu Val Trp Pro Pro Gln
20 25 30

Leu Gln Glu Gln Ala Pro Met Ala Gly Ala Leu Asn Arg Lys Glu Ser
35 40 45

Phe Leu Leu Leu Ser Leu His Asn Arg Leu Arg Ser Trp Val Gln Pro
50 55 60

Pro Ala Ala Asp Met Arg Arg Leu Asp Trp Ser Asp Ser Leu Ala Gln
65 70 75 80

Leu Ala Gln Ala Arg Ala Ala Leu Cys Gly Ile Pro Thr Pro Ser Leu
85 90 95

Ala Ser Gly Leu Trp Arg Thr Leu Gln Val Gly Trp Asn Met Gln Leu
100 105 110

Leu Pro Ala Gly Leu Ala Ser Phe Val Glu Val Val Ser Leu Trp Phe
115 120 125

Ala Glu Gly Gln Arg Tyr Ser His Ala Ala Gly Glu Cys Ala Arg Asn
130 135 140

Ala Thr Cys Thr His Tyr Thr Gln Leu Val Trp Ala Thr Ser Ser Gln
145 150 155 160

Leu Gly Cys Gly Arg His Leu Cys Ser Ala Gly Gln Ala Ala Ile Glu
165 170 175

Ala Phe Val Cys Ala Tyr Ser Pro Gly Gly Asn Trp Glu Val Asn Gly
180 185 190

Lys Thr Ile Ile Pro Tyr Lys Lys Gly Ala Trp Cys Ser Leu Cys Thr

195 200 205

Ala Ser Val Ser Gly Cys Phe Lys Ala Trp Asp His Ala Gly Gly Leu

210 215 220

Cys Glu Val Pro Arg Asn Pro Cys Arg Met Ser Cys Gln Asn His Gly

225 230 235 240

Arg Leu Asn Ile Ser Thr Cys His Cys His Cys Pro Pro Gly Tyr Thr

245 250 255

Gly Arg Tyr Cys Gln Val Arg Cys Ser Leu Gln Cys Val His Gly Arg

260 265 270

Phe Arg Glu Glu Glu Cys Ser Cys Val Cys Asp Ile Gly Tyr Gly Gly

275 280 285

Ala Gln Cys Ala Thr Lys Val His Phe Pro Phe His Thr Cys Asp Leu

290 295 300

Arg Ile Asp Gly Asp Cys Phe Met Val Ser Ser Glu Ala Asp Thr Tyr

305 310 315 320

Tyr Arg Ala Arg Met Lys Cys Gln Arg Lys Gly Gly Val Leu Ala Gln

325 330 335

Ile Lys Ser Gln Lys Val Gln Asp Ile Leu Ala Phe Tyr Leu Gly Arg

340 345 350

Leu Glu Thr Thr Asn Glu Val Ile Asp Ser Asp Phe Glu Thr Arg Asn

355 360 365

Phe Trp Ile Gly Leu Thr Tyr Lys Thr Ala Lys Asp Ser Phe Arg Trp

370 375 380

Ala Thr Gly Glu His Gln Ala Phe Thr Ser Phe Ala Phe Gly Gln Pro

385 390 395 400

Asp Asn His Gly Phe Gly Asn Cys Val Glu Leu Gln Ala Ser Ala Ala
405 410 415
Phe Asn Trp Asn Asn Gln Arg Cys Lys Thr Arg Asn Arg Tyr Ile Cys
420 425 430
Gln Phe Ala Gln Glu His Ile Ser Arg Trp Gly Pro Gly Ser
435 440 445

<210> 6
<211> 197
<212> PRT
<213> Homo sapiens
<400> 6

Met Pro Pro Ala Gly Leu Arg Arg Ala Ala Pro Leu Thr Ala Ile Ala
1 5 10 15
Leu Leu Val Leu Gly Ala Pro Leu Val Leu Ala Gly Glu Asp Cys Leu
20 25 30
Trp Tyr Leu Asp Arg Asn Gly Ser Trp His Pro Gly Phe Asn Cys Glu
35 40 45
Phe Phe Thr Phe Cys Cys Gly Thr Cys Tyr His Arg Tyr Cys Cys Arg
50 55 60
Asp Leu Thr Leu Leu Ile Thr Glu Arg Gln Gln Lys His Cys Leu Ala
65 70 75 80
Phe Ser Pro Lys Thr Ile Ala Gly Ile Ala Ser Ala Val Ile Leu Phe
85 90 95
Val Ala Val Val Ala Thr Thr Ile Cys Cys Phe Leu Cys Ser Cys Cys

tggcgcctcc tggctatgct ggccgggctc tacgccttct tcctgtttga gaacctcttc 1260

aatctcctgc tgcccaggga cccggaggac ctggaggacg ggccctgcgg ccacagcagc 1320

catagccacg ggggccacag ccacggtgtg tccctgcagc tggcacccag cgagctccgg 1380

cagcccaagc ccccccacga gggctcccgc gcagacctgg tggcggagga gagcccggag 1440

ctgctgaacc ctgagcccag gagactgagc ccagagttga ggctactgcc ctatatgatc 1500

actctgggcg acgccgtgca caacttcgcc gacgggctgg ccgtgggcgc cgccttcgcg 1560

tcctcctgga agaccgggct ggccacctcg ctggccgtgt tctgccacga gttgccacac 1620

gagctggggg acttcgccgc cttgctgcac gcggggctgt ccgtgcgcca agcactgctg 1680

ctgaacctgg cctccgcgct cacggccttc gctggtctct acgtggcact cgcggttgga 1740

gtcagcgagg agagcgaggc ctggatcctg gcagtggcca ccggcctgtt cctctacgta 1800

gcactctgcg acatgctccc ggcgatgttg aaagtacggg acccgcggcc ctggctcctc 1860

ttcctgctgc acaacgtggg cctgctgggc ggctggaccg tcctgctgct gctgtccctg 1920

tacgaggatg acatcacctt ctga 1944

<210> 15

<211> 1341

<212> DNA

<213> Homo sapiens

<400> 15

atgctgcatc cagagacctc ccctggccgg gggcatctcc tggctgtgct cctggccctc 60

cttggcaccg cctgggcaga ggtgtggcca ccccagctgc aggagcaggc tccgatggcc 120

ggagccctga acaggaagga gagtttcttg ctcctctccc tgcacaaccg cctgcgcagc 180

tgggtccagc cccctgcggc tgacatgcgg aggctggact ggagtgacag cctggcccaa 240

ctggctcaag ccagggcagc cctctgtgga atcccaaccc cgagcctggc gtccggcctg 300

tggcgcaccc tgcaagtggg ctggaacatg cagctgctgc ccgcgggctt ggcgtccttt 360

gttgaagtgg tcagcctatg gtttgcagag gggcagcggt acagccacgc ggcaggagag 420

tgtgctcgca acgccacctg cacccactac acgcagctcg tgtgggccac ctcaagccag 480

ctgggctgtg ggcggcacct gtgctctgca ggccaggcag cgatagaagc ctttgtctgt 540

gcctactccc ccggaggcaa ctgggaggtc aacgggaaga caatcatccc ctataagaag 600

ggtgcctggt gttcgctctg cacagccagt gtctcaggct gcttcaaagc ctgggaccat 660

gcaggggggc tctgtgaggt ccccaggaat ccttgtcgca tgagctgcca gaaccatgga 720

cgtctcaaca tcagcacctg ccactgccac tgtcccctg gctacacggg cagatactgc 780

caagtgaggt gcagcctgca gtgtgtgcac ggccggttcc gggaggagga gtgctcgtgc 840

gtctgtgaca tcggctacgg gggagcccag tgtgccacca aggtgcattt tcccttccac 900

acctgtgacc tgaggatcga cggagactgc ttcatggtgt cttcagaggc agacacctat 960

tacagagcca ggatgaaatg tcagaggaaa ggcggggtgc tggcccagat caagagccag 1020

aaagtgcagg acatcctcgc cttctatctg ggccgcctgg agaccaccaa cgaggtgatt 1080

gacagtgact tcgagaccag gaacttctgg atcgggctca cctacaagac cgccaaggac 1140

tccttccgct gggccacagg ggagcaccag gccttcacca gttttgcctt tgggcagcct 1200

gacaaccacg ggtttggcaa ctgcgtggag ctgcaggctt cagctgcctt caactggaac 1260

aaccagcgct gcaaaacccg aaaccgttac atctgccagt ttgcccagga gcacatctcc 1320

cggtggggcc cagggtcctg a 1341

<210> 16

<211> 594

<212> DNA

<213> Homo sapiens

<400> 16

atgccacccg cggggctccg ccgggccgcg ccgctcaccg caatcgctct gttggtgctg 60

ggggctcccc tggtgctggc cggcgaggac tgcctgtggt acctggaccg gaatggctcc 120

630 635 640

gag gat gac atc acc ttc tgataccctg ccctagtccc ccacctttga 2026

Glu Asp Asp Ile Thr Phe

645

cttaagatcc cacacctcac aaacctacag cccagaaacc agaagcccct atagaggccc 2086

cagtcccaac tccagtaaag acactcttgt ccttgg 2122

<210> 25

<211> 1775

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (62)..(1402)

<400> 25

aaaacaagcc gggtggctga gccaggctgt gcacggagtg cctgacgggc ccaacagacc 60

c atg ctg cat cca gag acc tcc cct ggc cgg ggg cat ctc ctg gct gtg 109

Met Leu His Pro Glu Thr Ser Pro Gly Arg Gly His Leu Leu Ala Val

1 5 10 15

ctc ctg gcc ctc ctt ggc acc gcc tgg gca gag gtg tgg cca ccc cag 157

Leu Leu Ala Leu Leu Gly Thr Ala Trp Ala Glu Val Trp Pro Pro Gln

20 25 30

ctg cag gag cag gct ccg atg gcc gga gcc ctg aac agg aag gag agt 205

Leu Gln Glu Gln Ala Pro Met Ala Gly Ala Leu Asn Arg Lys Glu Ser

35 40 45

ttc ttg ctc ctc tcc ctg cac aac cgc ctg cgc agc tgg gtc cag ccc 253
Phe Leu Leu Leu Ser Leu His Asn Arg Leu Arg Ser Trp Val Gln Pro
50 55 60

cct gcg gct gac atg cgg agg ctg gac tgg agt gac agc ctg gcc caa 301
Pro Ala Ala Asp Met Arg Arg Leu Asp Trp Ser Asp Ser Leu Ala Gln
65 70 75 80

ctg gct caa gcc agg gca gcc ctc tgt gga atc cca acc ccg agc ctg 349
Leu Ala Gln Ala Arg Ala Ala Leu Cys Gly Ile Pro Thr Pro Ser Leu
85 90 95

gcg tcc ggc ctg tgg cgc acc ctg caa gtg ggc tgg aac atg cag ctg 397
Ala Ser Gly Leu Trp Arg Thr Leu Gln Val Gly Trp Asn Met Gln Leu
100 105 110

ctg ccc gcg ggc ttg gcg tcc ttt gtt gaa gtg gtc agc cta tgg ttt 445
Leu Pro Ala Gly Leu Ala Ser Phe Val Glu Val Val Ser Leu Trp Phe
115 120 125

gca gag ggg cag cgg tac agc cac gcg gca gga gag tgt gct cgc aac 493
Ala Glu Gly Gln Arg Tyr Ser His Ala Ala Gly Glu Cys Ala Arg Asn
130 135 140

gcc acc tgc acc cac tac acg cag ctc gtg tgg gcc acc tca agc cag 541
Ala Thr Cys Thr His Tyr Thr Gln Leu Val Trp Ala Thr Ser Ser Gln
145 150 155 160

ctg ggc tgt ggg cgg cac ctg tgc tct gca ggc cag gca gcg ata gaa 589
Leu Gly Cys Gly Arg His Leu Cys Ser Ala Gly Gln Ala Ala Ile Glu
165 170 175

gcc ttt gtc tgt gcc tac tcc ccc gga ggc aac tgg gag gtc aac ggg 637

Ala Phe Val Cys Ala Tyr Ser Pro Gly Gly Asn Trp Glu Val Asn Gly
180 185 190

aag aca atc atc ccc tat aag aag ggt gcc tgg tgt tcg ctc tgc aca 685
Lys Thr Ile Ile Pro Tyr Lys Lys Gly Ala Trp Cys Ser Leu Cys Thr
195 200 205

gcc agt gtc tca ggc tgc ttc aaa gcc tgg gac cat gca ggg ggg ctc 733
Ala Ser Val Ser Gly Cys Phe Lys Ala Trp Asp His Ala Gly Gly Leu
210 215 220

tgt gag gtc ccc agg aat cct tgt cgc atg agc tgc cag aac cat gga 781
Cys Glu Val Pro Arg Asn Pro Cys Arg Met Ser Cys Gln Asn His Gly
225 230 235 240

cgt ctc aac atc agc acc tgc cac tgc cac tgt ccc cct ggc tac acg 829
Arg Leu Asn Ile Ser Thr Cys His Cys His Cys Pro Pro Gly Tyr Thr
245 250 255

ggc aga tac tgc caa gtg agg tgc agc ctg cag tgt gtg cac ggc cgg 877
Gly Arg Tyr Cys Gln Val Arg Cys Ser Leu Gln Cys Val His Gly Arg
260 265 270

ttc cgg gag gag gag tgc tcg tgc gtc tgt gac atc ggc tac ggg gga 925
Phe Arg Glu Glu Glu Cys Ser Cys Val Cys Asp Ile Gly Tyr Gly Gly
275 280 285

gcc cag tgt gcc acc aag gtg cat ttt ccc ttc cac acc tgt gac ctg 973
Ala Gln Cys Ala Thr Lys Val His Phe Pro Phe His Thr Cys Asp Leu
290 295 300

agg atc gac gga gac tgc ttc atg gtg tct tca gag gca gac acc tat 1021
Arg Ile Asp Gly Asp Cys Phe Met Val Ser Ser Glu Ala Asp Thr Tyr

305 310 315 320

tac aga gcc agg atg aaa tgt cag agg aaa ggc ggg gtg ctg gcc cag 1069
Tyr Arg Ala Arg Met Lys Cys Gln Arg Lys Gly Gly Val Leu Ala Gln

325 330 335

atc aag agc cag aaa gtg cag gac atc ctc gcc ttc tat ctg ggc cgc 1117
Ile Lys Ser Gln Lys Val Gln Asp Ile Leu Ala Phe Tyr Leu Gly Arg

340 345 350

ctg gag acc acc aac gag gtg att gac agt gac ttc gag acc agg aac 1165
Leu Glu Thr Thr Asn Glu Val Ile Asp Ser Asp Phe Glu Thr Arg Asn

355 360 365

ttc tgg atc ggg ctc acc tac aag acc gcc aag gac tcc ttc cgc tgg 1213
Phe Trp Ile Gly Leu Thr Tyr Lys Thr Ala Lys Asp Ser Phe Arg Trp

370 375 380

gcc aca ggg gag cac cag gcc ttc acc agt ttt gcc ttt ggg cag cct 1261
Ala Thr Gly Glu His Gln Ala Phe Thr Ser Phe Ala Phe Gly Gln Pro

385 390 395 400

gac aac cac ggg ttt ggc aac tgc gtg gag ctg cag gct tca gct gcc 1309
Asp Asn His Gly Phe Gly Asn Cys Val Glu Leu Gln Ala Ser Ala Ala

405 410 415

ttc aac tgg aac aac cag cgc tgc aaa acc cga aac cgt tac atc tgc 1357
Phe Asn Trp Asn Asn Gln Arg Cys Lys Thr Arg Asn Arg Tyr Ile Cys

420 425 430

cag ttt gcc cag gag cac atc tcc cgg tgg ggc cca ggg tcc 1399
Gln Phe Ala Gln Glu His Ile Ser Arg Trp Gly Pro Gly Ser

435 440 445

tgaggcctga ccacatggct ccctcgcctg ccctgggagc accggctctg cttacctgtc 1459

cgcccacctg tctggaacaa gggccaggtt aagaccacat gcctcatgtc caaagaggtc 1519

tcagaccttg cacaatgcca gaagttgggc agagagaggc agggaggcca gtgagggcca 1579

gggagtgagt gttagaagaa gctggggccc ttcgcctgct tttgattggg aagatgggct 1639

tcaattagat ggcaaaggag aggacaccgc cagtggtcca aaaaggctgc tctcttccac 1699

ctggcccaga ccctgtgggg cagcggagct tccctgtggc atgaacccca cagggtatta 1759

aattatgaat cagctg 1775

<210> 26

<211> 1372

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (327)..(920)

<400> 26

aactgcccgc agtgcccatg gtggctcgga tgggaggaac caccgcggag ccggggacag 60

ggggagcagg gcagtgctct gctgggtgag gggcacccag ctccagaggc taggtgggcg 120

tcgctggtgg gtggactcct gggcgctgcg cggagccgcg ccggctgggt tagcgcgggc 180

ggggcgctta gtcccacccc cagaggaggc ggaagaggag cccgagcctg gccgcgggct 240

gggccccgcc gcagctccag ctggccggct tggtcctgcg gtcccttctc tgggaggccc 300

gaccccggcc gcgcccagcc cccacc atg cca ccc gcg ggg ctc cgc cgg gcc 353

Met Pro Pro Ala Gly Leu Arg Arg Ala

1 5

gcg ccg ctc acc gca atc gct ctg ttg gtg ctg ggg gct ccc ctg gtg 401